Integrative genome-wide analysis reveals a robust genomic glioblastoma signature associated with copy number driving changes in gene expression.

Marie de Tayrac, Amandine Etcheverry, Marc Aubry, Stephan Saïkali, Abderrahmane Hamlat, Véronique Quillien, André Le Treut, Marie-Dominique Galibert, Jean Mosser

▶ **To cite this version:**

Marie de Tayrac, Amandine Etcheverry, Marc Aubry, Stephan Saïkali, Abderrahmane Hamlat, et al.. Integrative genome-wide analysis reveals a robust genomic glioblastoma signature associated with copy number driving changes in gene expression.. Genes, Chromosomes & Cancer, 2009, 48 (1), pp.55-68. 10.1002/gcc.20618 . inserm-00352960

HAL Id: inserm-00352960

https://www.hal.inserm.fr/inserm-00352960

Submitted on 17 Nov 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Integrative Genome-wide Analysis Reveals a Robust Genomic Glioblastoma Signature Associated with Copy Number Driving Changes in Gene Expression.

Marie de Tayrac[1,2], Amandine Etcheverry[3], Marc Aubry[3], Stephan Saïkali[4,5], Abderrahmane Hamlat[6], Veronique Quillien[1,7], André Le Treut[1,2], Marie-Dominique Galibert[1,2] and Jean Mosser[1,2,3] *

1 CNRS-UMR 6061, Regulation of transcription and oncogenesis, IFR140 GFAS, Faculty of Medicine, Rennes, France

2 Department of Biochemistry and Molecular Genetics, Medical Genomic Unit, CHU Rennes, France

3 Transcriptomic platform, OUEST-Genopole®, Rennes, France

4 Department of pathology, CHU Rennes, France

5 Biological Ressources Center, CHU Rennes, France

6 Department of Neurosurgery, CHU Rennes, France

7 CRLCC, Centre Eugene Marquis, Rennes, France

* Correspondence to:

Jean Mosser - jean.mosser@univ-rennes1.fr tel: +33 223 234 951 fax: +33 223 234 478

Faculté de Médecine, 2 Avenue du Professeur Léon Bernard

35043 Rennes, France

Abbreviated title: Integrative genome-wide analysis of Glioblastoma

Supported by: This work was supported in part by the Ligue contre le Cancer foundation, the Région Bretagne (PRIR) and the Cancéropôle Grand-Ouest consortium 'From the core to beyond the margin'.

Abstract

Glioblastoma multiforme shows multiple chromosomal aberrations, the impact of which on gene expression remains unclear. To investigate this relationship and to identify putative initiating genomic events, we integrated a paired copy number and gene expression survey in glioblastoma using whole human genome arrays. Loci of recurrent copy number alterations were combined with gene expression profiles obtained on the same tumor samples. We identified a set of 406 '*cis*-acting DNA targeted genes' corresponding to genomic aberrations with direct copy-number-driving changes in gene expression, defined as genes with either significantly concordant or correlated changes in DNA copy number and expression. Functional annotation revealed that these genes participate in key processes of cancer cell biology, providing insights into the genetic mechanisms driving glioblastoma. The robustness of the gene selection was validated on an external microarray data set including 81 glioblastomas and 23 non-neoplastic brain samples. The integration of array CGH and gene expression data highlights a robust '*cis*-acting DNA targeted genes' signature that may be critical for glioblastoma progression, with two tumor suppressor genes *PCDH9* and *STARD13* that could be involved in tumor invasiveness and resistance to etoposide.

INTRODUCTION

Glioblastoma, the most devastating of the primary brain tumors, is characterized by deregulation of multiple pathways, such as EGFR/PTEN/Akt/mTOR, TP53/MDM2/$p14^{ARF}$ and $p16^{INK4a}$/RB. Several signaling molecules in such cascades are already the targets of therapies for glioblastomas but improvements remain modest from a clinical standpoint (Stupp et al., 2007). Identification of further tumor biomarkers is thus needed to provide new molecular targeted therapies. Important insights into tumor suppressor genes and oncogenes will probably be provided by identifying genomic aberrations inducing direct changes in gene expression, i.e., genes with expression levels either significantly concordant or correlated with changes in DNA copy number.

DNA copy number alterations (CNAs) are generally more numerous in malignant tumors than in benign ones, and could be classified as both causal and random genetic events. Some CNAs have a direct effect on gene expression and are likely to be more critical than others in the biology of cancer. Such CNAs can result either in the loss of tumor suppressor gene function or in the over-expression and activation of oncogenes. Both mechanisms constitute putative early oncogenic steps. A better understanding of the initiating molecular determinants of malignant tumors will require the identification of the CNAs that are functionally significant.

This task is particularly challenging for glioblastomas because of their highly rearranged genome (Bredel et al., 2005 a, Maher et al., 2006), and by the large number of genes that have been implicated at the transcriptome level (Nutt et al., 2003, Tso et al., 2006).

Recently, some genome-scale studies of glioblastoma described the relationship between DNA dosage and gene expression (Nigro et al., 2005, Liu et al., 2006, Phillips et al., 2006), but with some weaknesses: (1) intra-tumor variability was not taken into account; (2) direct gene level correlations were not really possible due to the use of low resolution arrays; and most importantly (3) genome and transcriptome data sets were unpaired, prone to provide false positives, and so to miss targeted genes.

In this study, we performed paired genome-wide analyses of glioblastoma, focusing on genes that showed concordant CNAs and expression patterns. High-resolution maps of chromosomal alterations were obtained by performing array-based comparative genomic hybridizations (arrays CGH) on 19 glioblastomas. Gene expression profiling was carried out on the same tumor samples, and compared to those obtained on non-neoplastic brain samples. We validated and investigated our result in an independent publicly available microarray data set of 81 glioblastomas and 23 normal brains. The associated annotation and network analyses provide insights into the genetic mechanisms driving glioblastoma.

MATERIALS AND METHODS

Tissue Samples

A total of 76 glioblastomas and 8 normal brain samples were used in this study. Fresh-frozen glioblastoma samples from patients admitted to the Neurosurgery Departments of Brittany University Hospitals (Rennes and Brest) were collected with informed consent and subjected to WHO classification. Histology was confirmed by hematoxylin-eosin staining of paraffin-embedded blocks. The non-neoplastic brain tissues were obtained from normal white matter area taken from patients undergoing surgery for chronic epilepsy. For microarray and array CGH hybridization, a set of 19 glioblastoma and 4 normal brain samples were used. Blood was available for all these patients. Each snap-frozen tumor block was cryodissected in 10µm sections. The first section as well as sections obtained every 100µm were stained to select tissue with at least 70 percent of tumor cells and to exclude necrotic areas and widespread blood vessels. To allow a paired and accurate comparison between closely related biological materials, the adjacent sections were alternatively pooled in different tubes for RNA and DNA extraction. For subsequent real-time reverse transcription-PCR validation of selected genes, we used an independent set of 57 glioblastoma samples.

Nucleic Acid Preparation

Tumor DNA was extracted according to manufacturer's instructions (NucleoSpin Tissue Kit, Macherey Nagel, Düren, DE). Blood DNA (reference) was isolated from peripheral blood leucocytes using a classical saline extraction. Total RNA was isolated using Macherey-Nagel NucleoSpin RNAII Kit. RNA integrity was confirmed using the Agilent

2100 Bioanalyzer (Agilent, Palo Alto, CA).

Arrays CGH Profiling

Genomic analyses were performed on Human 44K Agilent arrays CGH (Agilent Technologies, Santa Clara, CA) according to the manufacturer's protocol. Briefly, 5µg of DNA were double-digested (*AluI* and *RsaI*, Promega, Madison, WI, USA). Tumor DNA and reference DNA were both labeled by random priming with Cy3-dCTP and Cy5-dCTP for dye-swap experimental design. Tumor and reference DNA were pooled and hybridized (65°C, 48h). Arrays were washed and scanned on an Agilent G2565BA microarray scanner. Data were extracted and flagged with the Feature Extraction software (FE v9.4.1, CGH_44k_1005 protocol). Data preprocessing was carried out using limma (R package) from Bioconductor (Smyth et al., 2005). Values were median normalized and fluorescence $\log_2$ ratios were calculated as the average of two-paired arrays (dye-swap) except for one pair of arrays from which only one array met the quality criteria. Missing values were imputed with the k-nearest neighbors method implemented in impute (R package).

Array CGH Data Analysis

Log_2 ratios (tumor *vs.* reference) of probe intensities were plotted according to their genomic position, chromosome by chromosome. DNA copy number alterations were identified using the Gain and Loss Analysis of DNA algorithm (Hupé et al., 2004) implemented in GLAD (R package). This method uses Adaptive Weights Smoothing (AWS) procedure to detect breakpoints from array CGH profiles, and assigns a copy number status to each altered or normal chromosomal region. A 'segmented' data set

was generated by determining uniform copy number segment boundaries and by replacing normalized $\log_2$ ratio for each probe by the calculated smoothing values. As done in the multiple myeloma study of Aguirre et al. (2004), distribution of the smoothing values was used to define thresholds for the analysis of hemizygous deletions, homozygous deletions, gains and amplifications. Thresholds for low-copy number gain and hemizygous deletion were set at -0.15 and +0.15 respectively (±6 SD of the middle 75% of the data). Thresholds for high amplitude events were chosen at +0.9 for amplification and at -0.40 for homozygous deletion. Minimal common regions (MCRs) were defined as loci with CNAs in at least two samples; one of them showing an extreme CNA event defined by thresholds +0.29 and -0.29, <99% and <1% quantiles. Unsupervised hierarchical clustering was done on the 'segmented' data matrix using the Support Tree option in MeV (http://www.tigr.org/software/tm4/midas.html). Consensus clusters (average linkage and Pearson correlation metric) were built by bootstrap.

Gene Expression Profiling

Gene expression profiling was performed using the Agilent Whole Human oligo-Microarray Kit 4x44K multiplex format (Agilent Technologies, Santa Clara, CA, USA), with manufacturer's recommended procedures for microarray-based one-color (http://www.chem.agilent.com/temp/rad37FF4/00064034.pdf). Briefly, 350ng of total RNA with control RNA Spike In were amplified and labeled with Cy3-CTP (Agilent Technologies, Santa Clara, CA). 1.65µg of Cy3-labeled RNA were hybridized (65°C, 17h) per array. The processed Multiplicative Detrend FE data (FE v9.4.1, GE1-v5_91_0806 protocol) of scanned images were median normalized and missing values were imputed (limma and impute, R packages).

Expression Data Analysis

Analysis of gene expression was conducted to highlight the genes differentially expressed between glioblastomas and normal brain tissues. The significance in differential gene expressions was determined using standard Student's *t*-test of two groups. To account for multiple hypothesis testing, we computed adjusted *p*-values by controlling the false discovery rate (FDR) with the Benjamini & Hochberg (BH) procedure implemented in multtest (R package). Differentially expressed genes were defined as follows: BH adjusted *p*-value<0.01 and absolute mean $\log_2$ ratio (glioblastoma *vs.* mean normal brain) greater than 2.

The complete dataset has been submitted to the Gene Expression Omnibus Data (GEO) public database at NCBI, and the accession number is GSE10878.

Integrated Copy Number and Expression Analyses

Combination of genome and transcriptome datasets was done gene-by-gene for all the annotated genes that were present on both arrays. We used two approaches to identify all the genes whose transcription levels were potentially affected by DNA alterations. A schematic of our approach is provided in Figure 1. In the *Targeted study* (Figure 1 A), we identified the genes differentially expressed between glioblastoma and normal brain (BH *p*-value<0.05), located in MCR and concordant to the corresponding CNA, e.g. we selected those that were over-expressed or under-expressed and located in a region of gain or loss respectively. In the *Correlation study* (Figure 1 B), we evaluated the direct influence of copy number alterations on gene expression in MCR. We identified the

genes with highly correlated DNA 'segmented' values and expression patterns (Pearson's correlation coefficient up to 0.7).

Gene Ontology, Canonical Pathway, and Functional Network Analyses

Functional annotation analyses were performed with the NIH-DAVID software (version 2.1b, http://david.abcc.ncifcrf.gov/) and the method developed by Aubry et al. (2006). We used the first method with the parameters: GOTERM_BP_ALL, KEGG_PATHWAY and SP_PIR_Keywords; the significance threshold was set on a *p*-value<0.05 with Benjamini multiple testing correction. The second one was used to provide deeper informative annotations by combining evidence and literature with Gene Ontology Annotation database and PubGene biomedical literature index. Functional network analyses were executed using the Web-delivered application from Ingenuity Pathways Analysis[4] that enables the visualization and exploration of molecular interaction networks in gene expression data.

Real-time quantitative PCR (Q-PCR) for PCDH9 and STARD13

Q-PCR reactions were done with the 7900HT Fast Real-Time PCR System using the SYBR™Green PCR Master Mix (Applied Biosystems®). *B2M*, β-2 micoglobulin, RNA was chosen as internal control. Calibration was done with FirstChoice® Human Brain Reference Total RNA (Applied Biosystems®). The relative amounts of the gene transcripts were determined using the ΔΔCt method, as described by the manufacturer. The following forward (F) and reverse (R) primers were designed using Primer3 (v.0.4.0): *B2M*, F: 5'-TCCAACATCAACATCTTGGT-3' and R: 5'-TCCCCCAAATTCTAAGCAGA-3'; *PCDH9*, F: 5'-GCATATTGTCACTTAGGTCAAACCA-3' and R: 5'-

GTCATGCCTTAACAAAAACCTCCT-3'; *STARD13*, F: 5'-TGCTAATGGATCGAATGTGCTT-3' and R: 5'-TTCTCCAACACCAGTTGCTAAATC-3'.

Comparison with published expression data

We evaluated the robustness of gene selections on a publicly available microarray data set. The expression data from Sun et al. (2006) (GDS1962) were used to generate the comparison data set of gene expression changes between glioblastoma (n = 81 samples) and normal brain (n = 23 samples). The GDS1962 was downloaded from the Gene Expression Omnibus database[5] and managed with GEOquery (R package). A data matrix was generated by R programming, with the mean centered and scaled values corresponding to the processed data per array. We downloaded the annotation of the Affymetrix U133 Plus 2.0 Array platform from the Ensembl website. We retrieved the genes we were interested in by identifying probe sets with gene symbols. For heatmap representation, values were expressed as $\log_2$ ratio of glioblastoma *vs.* mean of all normal brain samples.

RESULTS

Recurrent and Novel Genomic Changes in Glioblastoma

Glioblastomas (n=19) were analyzed by array CGH in a dye-swap experiment design to support the normalization step and to provide a robust evaluation of the genomic profiles. Only somatic changes were defined since each tumor DNA was hybridized with the corresponding patient blood DNA. Segmentation analysis identified large aberrations at the genome level as well as focal higher-amplitude recurrent CNAs (MCRs). A summary of the CNAs is shown in Figure 2 A. The most frequent imbalances were: gain of chromosomes 7 (73%, gain of both 7p and 7q in 47%, 7p alone 10%, and 7q alone in 16%) and 20 (16%); and loss of 9p (58%), chromosome 10 (58%), parts of chromosome 13 (31%), and 22q (21%).

The MCRs (2285 DNA segments and 4816 genes) corresponding to either amplification (log2 ratio > 0.9) or homozygous deletion (log2 ratio < -0.40) are presented in Supplementary Table 1 (A and B). These high-amplitude events span a median size of 8.4Mb with an average of 11 known genes. The following previously well-characterized amplicons were identified: co-amplification of *PDGFRA*, *KIT* and *KDR* (three tumors), co-amplification of *EGFR*, *SEC61G* (six cases), co-amplification of *CDK4* and *MDM2* (one tumor) and amplicon at 1q32.1 (two cases) including a gene encoding a catalytic subunit of Pi3k (*PIK3C2B*) (Knobbe et al., 2003). The well-known deletion of *CDKN2A* and *CDKN2B* with co-deletion of the putative tumor suppressor gene *MTAP* (Schmid et al., 2000) was found in four cases. The recently identified glioblastoma deletion of tumor-suppressor gene *CDKN2C* was also present in one tumor (Solomon et al., 2008). In

addition, we identified the following new amplicons: at 20q11-13, including *POFUT1* known to be essential for Notch function (Kroes et al., 2007), at 9p22.1, including *DNAJA1* that encodes a co-chaperone of heat shock protein 70 (Wang et al., 2006) and at 11p13, including *PAX6* (Daugaard et al., 2007). New homozygous deletions included particularly DKK1 (10q11.2) that is a Wnt/β-catenin pathway inhibitor shown to be pro-apoptotic in brain tumor cells (Shou et al., 2002) and other genes (*CNTNAP3*, 9p13.1; *GLUD2*, in Xq24, and, *BAGE* and *BAGE4*, in 21p11.1).

MCRs Cluster Glioblastoma Around EGFR and STARD13 Status

To probe the organization of MCRs across the tumor set, we performed unsupervised hierarchical clustering of glioblastomas in the space of MCRs (without taking into account the sex chromosomes). Cluster analysis (Figure 2 B) emphasized common recurrent aberrations, including the gain of whole chromosome 7 and losses of chromosome 10 and 9p. Notably, cluster analysis highlighted two sample groups depending apparently on whether *EGFR* (7p11.2) was amplified or not (subtype-1 and subtype-2, respectively). In addition, it showed that three glioblastomas (one with *EGFR* amplification, the others without gain of chromosome 7) had particular genomic profiles (see Supplementary Table 2). For these three patients, neither re-analysis of the histological nor clinical data was able to phenotypically appreciate these particularities. In order to more directly assess the relation between MCRs and the two groups (subtype-1 and subtype-2), we applied a supervised analysis using a standard Student's *t*-test. Supervised analysis showed that the two subtypes exhibited distinct patterns of MCRs (Supplementary Table 3). In particular, it confirmed that all the subtype-1-glioblastomas but none of the subtype-2-glioblastomas carried *EGFR* amplification at

7p11.2 locus (p=0.006). It also showed that: (i) subtype-1 presented more frequent hemizygous deletions at 13q31-13q34 (p=0.01) and 13q12-13q21 (p=0.03); (ii) gains at 7p21.3-p21.2, 7q21.11, 7q21.2 and 7q31.1-q34 were more frequent (p < 0.01) in subtype-2 (77-88%) compared to subtype-1 (28-42%). We also investigated these differences at the transcriptome level (Student's *t*-test with BH correction, comparing expression values between subtype-1 and subtype-2 for genes located in the aberrant regions on chromosome 7 and 13q). Only two differential gene expressions were highlighted: *EGFR* over-expression (p=0.011) and *STARD13* under-expression (p=0.036).

Glioblastoma Expression Profiling Identifies Huge Amount of Alterations

Evaluation of genes that are differentially expressed in glioblastoma versus normal brain was undertaken using a standard Student *t*-test with BH correction (p<0.01) and absolute average $\log_2$ ratio greater than 2. Expression profiling identified 664 over-expressed genes in glioblastoma, including 56 genes over-expressed greater than 30-fold, and 1224 under-expressed genes, including157 genes under-expressed greater than 30-fold. A list of all the identified genes is provided in Supplementary File 1. From the top ten of the over-expressed genes we identified the antiapoptotic gene *BIRC5*/survivin (Blum et al., 2006) and the transcription factor *E2F2* (Okamoto et al., 2007), the activities of which have been already linked to glioblastoma. We also found two highly expressed mitotic kinases, *PBK* and *BUB1*, and a potential cell cycle regulator, *DLG7*/*HURP*, that have not been previously reported in glioblastoma. Proliferation-related genes were also included among the highly expressed genes. Some of them were previously well described as participating in glioblastoma progression, such

as *UBE2C* (Bredel et al., 2005 b), *MKI67* (Raghavan et al., 1990), *TOP2A* (van den Boom et al., 2002), *TNC* (Sarkar et al., 2006), *CHI3L1*/*YKL-40* (Tanwar et al., 2002), *MELK* (Liu et al., 2006) and *CD44* (Ylagan et al., 1997). The others, *NCAPG*, *KIF20A*, *CENPA* and *RRM2*, are novel glioblastoma-associated genes with reported functional roles in cytokinesis and/or cell proliferation (Geiman et al., 2004, Wonsey et al., 2005). Regarding the under-expressed genes, we identified some with known tumor suppressor functions in glioma (*CHD5* and *LGI1*). *CHD5* was identified last year as a tumor-suppressor that controls proliferation, apoptosis, and senescence via the p19(Arf)/p53 pathway in glioma (Bagchi et al., 2007). It has been suggested that the leucine-rich, glioma-inactivated gene 1 (*LGI1*) gene is a candidate tumor suppressor gene involved in progression of glial tumors (Chernova et al., 1998). In addition, four components of the Wnt/β-catenin signaling pathway were highly under-expressed: two Wnt antagonists, *WIF1* and *SFRP1*, and *PPP2R2C* (PP2A) and *WNT10B* (Kirikoshi et al., 2002). Such disruptions may result in an improper function of the Wnt/β-signaling pathway leading to aberrant cell proliferation and therefore explaining part of the glioblastoma progression.

Functional annotation of the highly differentially expressed genes (greater or less than 30-fold) underlined distinct biological processes according to groups of over- or under-expression. Highly over-expressed genes were significantly associated with the regulation of the mitotic cell cycle, and more precisely with the following GOTERM_BP_ALL: M phase (p=5.7e-7), microtubule-based process (p=3.9e-3) and sister chromatid segregation (p=0.01). Highly under-expressed genes were significantly associated with cell communication (p=1.8e-3), cell-cell signaling (p=3.1e-9) and neurophysiological process (p=4.9e-5).

Identification of the glioblastoma DNA targeted genes

To identify targeted genes, we used the paired array CGH and transcriptome data sets, measured on the same glioblastomas and restricted to MCRs. We combined these two paired data sets to focus only on the genes for which the expression was affected by aberrant copy number variations. Two possibilities were evaluated: genes with expression levels significantly concordant with changes in DNA copy number (*Targeted study*) and genes with expression levels directly correlated with such changes but not necessarily with a significant differential expression in all glioblastomas (*Correlated study*) (see Materials and Methods, Figure 1, for schematic representation).

The *Targeted study* showed that DNA copy number influenced gene expression across a 13.8% range of MCRs, corresponding to 261 Significant Targeted Genes (STGs) that were differentially expressed and concordant with MCR patterns. Representation of DNA aberrations and corresponding gene expression on two mirrored heatmaps clearly shows the unbalanced distribution of concordant over- and under-expressed genes on chromosomes (Figure 3 A). The 95 STGs found to be highly expressed in gain of regions were located in chromosome 1 (1p32.1), 4 (4q11-4q12), 7, 12 (12p11) and X (Xq24) and the 166 STGs under-expressed in deleted regions were located in 9p, chromosome 10 and 13q. Functional annotation analyses of the STGs were performed by taking into account under- and over-expressed groups separately. Significant enrichments were found with DAVID: the over-expressed STGs were particularly related to developmental processes (GOTERM_BP_ALL: development, morphogenesis and SP_PIR_KEYWORDS: developmental protein) and the under-expressed STGs to mRNA splicing (SP_PIR_KEYWORDS: alternative splicing). The method developed by Aubry et al. (2006) provided more details on tumorigenesis-related

processes linked to STGs including cell cycle (mitosis and apoptosis), cell adhesion, DNA repair and angiogenesis (Figure 3 B). To understand how STGs are related, they were replaced in pathways and molecular interactions established on the Ingenuity knowledge base. Networks and associated functions and disease were scored and checked for significance. The most relevant one was built around *EGFR* and displayed high-level functions in cancer and neurological disease. It contained 27 STGs (Figure 3 C), including the GBM-related genes *CAV1*, *DMBT1*, *EGFR*, *KDR* and *MGMT*. The over-expressed STGs were notably linked to cell proliferation and movement.

The *Correlated study* focused on the DNA alterations affecting gene expression on a distinct, individual and isolated manner. To achieve this, we performed direct correlation analyses of copy number and expression data on a gene-by-gene basis throughout the genome. Transcription levels were highly correlated to genomic patterns for 159 Correlated Genes (CGs), representing only 4.1% of the genes located in MCRs. The CGs were located on chromosomes 1, 4, 7, 9, 10 and 13, as for the STGs, but also on chromosomes 17, 20 and 22. The strong influence of DNA copy number on the CGs expression is evident by examination of the heatmap representations of CGs data for DNA aberrations and gene expression (Figure 4 A). The overall patterns of gene amplification and increased gene expression are concordant, i.e., a significant fraction of the highly amplified genes appear to be correspondingly highly expressed. Such concordance is also found for part of the genes located in losses. Functional annotation of CGs highlighted Gene Ontology terms (biological processes) mostly associated with cell cycle, DNA repair, RNA processing, and brain development (Figure 4 B). The top-scoring networks (Figure 4 C), built around *EGFR* and *PDGF*, contained glioma-related genes (*CDKN2A*, *CDKN2B*, *EGFR* and *MLL3*), genes that play a role in hematological

disorder (*CDKN2A*, *CDKN2B*, *EGFR*, *GIT1*, *MPDZ*, *NT5C3*, *OPRS1*, *PDGFA*, *PLCG1*, *RALA*, *SLC25A13*, *SRPK2* and *TIMM23)* and genes involved in cell division process (*CDKN2A*, *CDKN2B*, *DMTF1*, *PLCG1* and *ZC3HC1)*.

A Targeted Genes Signature for Glioblastoma

Following the results of both *Targeted* and *Correlated* analyses, we determined a set of 406 genes constituting the glioblastoma targeted genes signature. Fourteen genes were identified in both analyses. To evaluate the relevance of this glioblastoma genomic signature, we used a publicly available microarray data set (GDS1962) of 81 glioblastomas and 23 normal brain samples, from the study by Sun et al. (2006). We were able to map 369 of our 406 selected genes on the Affymetrix U133 Plus 2.0 Array platform, and generated the corresponding data matrix. We performed a Student *t*-test on this data table to identify the genes that were differentially expressed between glioblastoma and normal brain. With a risk level (BH) of 5%, we confirmed that 86% of the 369 genes were differentially expressed in glioblastoma (92% for the *Targeted study* and 73% for the *Correlated study*). More details are available in Supplementary File 2. A heatmap representation of all the glioblastomas (81 from the Sun et al. study [external] and 19 from the present study [local]) is provided in Figure 5 A. The strong similarity of the red:green profiles between the two data sets (external and local) illustrate the strong robustness of the targeted genes signature. Network analysis provided two top scoring networks: the first one (Figure 5 B) was built around *Rb* and *Nfκ**B* and the second (Figure 5 C) around *Akt*, *EGFR* and *PDGFR*. Overlaid functions ('proliferation of cells', 'hematological disorder' and 'developmental process of tumor cell lines') were mostly associated to over-expressed genes. Both networks included a substantial number of

genes that have been implicated in glioblastomagenesis.

To go further in the validation of the glioblastoma targeted genes signature, we probed separately the organization of the *Targeted* genes across the tumor set from the *Correlated* genes (Figure 6). We performed unsupervised hierarchical clustering of all the tumors in the space of each set of genes. The *Targeted* cluster analysis (Figure 6 A) defined two groups of tumors. However, the genes signature did not provide a clear distinction between the two groups, illustrating the homogeneous expression of the *Targeted* genes in glioblastoma. The Correlated cluster analysis (Figure 6 B) emphasized the heterogeneous part of the signature, also grouping glioblastomas into two subtypes, mainly depending on whether genes mapping to 7p11.2 were highly over-expressed or not and whether *CDKN2A* and *CDKN2B* were deleted or not.

PCDH9 and *STARD13* were considered interesting due to their biological function and because their under-expression was strongly correlated to the corresponding genomic state. *PCDH9* and *STARD13* under-expression was validated by RT Q-PCR on an independent set of 57 glioblastoma samples and on the same panel of glioblastomas (Figure 7). It was also confirmed by the GDS1962 analysis.

DISCUSSION

Glioblastomas show numerous DNA alterations of varying size and location, as well as large transcriptome modifications. The functional annotation analyses suggest that the molecular alterations occurring in glioblastoma promote cell growth, proliferation, survival and apoptotic resistance. However, identifying high-priority biomarkers by this way is still challenging. Consequently, we sought to identify targeted genes by integrating data on the genome and transcriptome levels. To achieve this task, we implemented a dual strategy that delineates genes with either *significantly concordant* or *correlated* changes in expression and in copy number.

A major advantage of choosing a dual strategy is that it bypasses limitations due to a small number of cases and due to inter-tumor heterogeneity. Indeed, a great part of the genes with copy number driving changes in expression are likely to be functionally significant even if occurring in only few cases. In our study, such genes were surveyed by the *Correlation study*. This is well illustrated in our analysis of *CDKN2A* and *CDKN2B*. These two genes are altered at the DNA level with an impact on transcription. The *Correlation study* but not the *Targeted* one identified these two genes, which were also validated by external data analysis. Moreover, the *Correlation study* and the validation step help to balance the lack of statistical power due to our small sample size cohort.

The dual strategy delineated two sets of genes that appear to have distinct biological relevance. The genes identified by the *Targeted study* had homogeneous transcriptome

modifications in glioblastomas. Even though gene expression was affected in almost all tumors, it was not related to CNAs in all glioblastomas. The homogeneous transcriptome signature must involve other regulatory mechanisms such as regulations by transcription factor, miRNA and DNA methylation. These widespread homogeneous changes are related to mitotic defects, chromosome segregation errors, disruption of cell adhesion processes, activation of DNA repair, escape from apoptosis and angiogenesis. Moreover, our network analysis revealed a top-scoring pathway potentially affecting glioblastoma cell invasion. It is therefore tempting to assume that the *Targeted genes* correspond to changes highly essential for glioblastoma development and that they are the most likely candidate oncogenes and tumor-suppressor genes linked to tumorigenesis.

In contrast, the genes identified by the *Correlation study* have an expression directly and highly linked to the DNA copy number status, consequently representing more sporadic events. The two clusters of glioblastoma suggest distinct pathways leading to glioblastoma development, raising the question whether these different sets of genes combinations result from compensatory functions within genetic pathways or whether they represent different clinical subclasses of glioblastoma. Our functional annotation and network analyses provided some information, suggesting that part of the heterogeneity was due to genes associated general cancer biology, as well as to genes that control normal brain.

We also identified *PCDH9* and *STARD13* as potential tumor suppressor genes in glioblastoma. *PCDH9* is a protocadherin predominantly expressed in the nervous system

and is highly similar to cadherin-related tumor suppressor precursor[6]. Three other protocadherins, *PCDH8*, *PCDH20*, *PCDH21*, were shown to be under-expressed in the *Targeted study*; *PCHD20* is a candidate tumor suppressor in non-small-cell lung cancers (Imoto et al., 2006). As these genes are likely to be involved in cell-cell interactions in the glial cell compartment, it will be of interest to investigate the links between their low expression levels, adhesion of glial cells and the promoted invasive growth of glioblastoma. StAR-related lipid transfer (START) domain containing 13 (*STARD13*) belongs, with *DLC1*, to the family of RhoGap proteins with START domains (Soccio et al., 2003). The corresponding genes are frequently lost in hepatocellular and breast carcinomas (Ching et al., 2003, Nagaraja et al., 2004). Both *STARD13* and *DLC1* are suggested to be candidate tumor suppressor genes (Yuan et al., 1998, Ng et al., 2006). Moreover, Hatch et al. have recently suggested that impairments of *STARD13* gene can participate in etoposide resistance via its role in ceramide signaling to the RhoA pathway (Hatch et al., 2007). A temozolomide/etoposide chemotherapy regime is currently being evaluated in trials for glioblastoma (Korones et al., 2003), and it will be of great interest to study response to this treatment with regard to presence or absence of *STARD13* deletions.

ACKNOWLEDGEMENTS

We would like to thank Cyrille Surbled for technical assistance and members of the CNRS-UMR6061 for discussions and comments on the manuscript. We acknowledge the Departments of Neurosurgery (CHU-Rennes, CHU-Brest) for contributing tumor samples. We also thank Phillip Jordan for his help in writing this manuscript.

REFERENCES

Aguirre AJ, Brennan C, Bailey G, Sinha R, Feng B, Leo C, Zhang Y, Zhang J, Gans JD, Bardeesy N, Cauwels C, Cordon-Cardo C, Redston MS, DePinho RA, Chin L. 2004. High-resolution characterization of the pancreatic adenocarcinoma genome.Proc Natl Acad Sci U S A 101:9067-9072.

Aubry M, Monnier A, Chicault C, de Tayrac M, Galibert M-D, Burgun A, Mosser J. 2006. Combining evidence, biomedical literature and statistical dependence: new insights for functional annotation of gene sets. BMC Bioinformatics 7:241.

Bagchi A, Papazoglu C, Wu Y, Capurso D, Brodt M, Francis D, Bredel M, Vogel H, Mills AA. 2007. CHD5 is a tumor suppressor at human 1p36. Cell 128:459-475.

Blum R, Jacob-Hirsch J, Rechavi G, Kloog Y. 2006. Suppression of survivin expression in glioblastoma cells by the Ras inhibitor farnesylthiosalicylic acid promotes caspase-dependent apoptosis. Mol Cancer Ther 5:2337-2347.

Bredel M, Bredel C, Juric D, Harsh GR, Vogel H, Recht LD, Sikic BI. 2005a. Functional network analysis reveals extended gliomagenesis pathway maps and three novel MYC-interacting genes in human gliomas. Cancer Res 65:8679-8689.

Bredel M, Bredel C, Juric D, Harsh GR, Vogel H, Recht LD, Sikic BI. 2005b. High-resolution genome-wide mapping of genetic alterations in human glial brain tumors. Cancer Res 65:4088-4096.

Chernova OB, Somerville RP, Cowell JK. 1998. A novel gene, LGI1, from 10q24 is rearranged and downregulated in malignant brain tumors. Oncogene 17:2873-2881.

Ching Y-P, Wong C-M, Chan S-F, Leung TH-Y, Ng DC-H, Jin D-Y, Ng IO-l. 2003. Deleted in liver cancer (DLC) 2 encodes a RhoGAP protein with growth suppressor

function and is underexpressed in hepatocellular carcinoma. J Biol Chem 278:10824-10830.

Daugaard M, Kirkegaard-Sørensen T, Ostenfeld MS, Aaboe M, Høyer-Hansen M, Orntoft TF, Rohde M, J\"a ataM. 2007. Lens epithelium-derived growth factor is an Hsp70-2 regulated guardian of lysosomal stability in human cancer. Cancer Res 67:2559-2567.

Geiman TM, Sankpal UT, Robertson AK, Chen Y, Mazumdar M, Heale JT, Schmiesing JA, Kim W, Yokomori K, Zhao Y and others. 2004. Isolation and characterization of a novel DNA methyltransferase complex linking DNMT3B with components of the mitotic chromosome condensation machinery. Nucleic Acids Res 32:2716-2729.

Hatch GM, Gu Y, Xu FY, Cizeau J, Neumann S, Park J-S, Loewen S, Mowat MRA. 2007. StARD13(Dlc-2) RhoGap Mediates Ceramide Activation of Phosphatidylglycerolphosphate Synthase and Drug Response in Chinese Hamster Ovary Cells. Mol Biol Cell.

Hupé PH, Stransky N, Thiery J-P, ois Radvanyi Fc, Barillot E. 2004. Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. Bioinformatics 20:3413-22.

Imoto I, Izumi H, Yokoi S, Hosoda H, Shibata T, Hosoda F, Ohki M, Hirohashi S, Inazawa J. 2006. Frequent silencing of the candidate tumor suppressor PCDH20 by epigenetic mechanism in non-small-cell lung cancers. Cancer Res 66:4617-4626.

Kirikoshi H, Katoh M. 2002. Expression and regulation of WNT10B in human cancer: up-regulation of WNT10B in MCF-7 cells by beta-estradiol and down-regulation of WNT10B in NT2 cells by retinoic acid. Int J Mol Med 10:507-511.

Knobbe CB, Reifenberger G. 2003. Genetic alterations and aberrant expression of genes related to the phosphatidyl-inositol-3'-kinase/protein kinase B (Akt) signal transduction pathway in glioblastomas. Brain Pathol 13:507-518.

Korones DN, Benita-Weiss M, Coyle TE, Mechtler L, Bushunow P, Evans B, Reardon DA, Quinn JA, Friedman H. 2003. Phase I study of temozolomide and escalating doses of oral etoposide for adults with recurrent malignant glioma. Cancer 97:1963-1968.

Kroes RA, Dawson G, Moskal JR. 2007. Focused microarray analysis of glyco-gene expression in human glioblastomas. J Neurochem 103 Suppl 1:14-24.

Liu F, Park PJ, Lai W, Maher E, Chakravarti A, Durso L, Jiang X, Yu Y, Brosius A, Thomas M and others. 2006a. A genome-wide screen reveals functional gene clusters in the cancer genome and identifies EphA2 as a mitogen in glioblastoma. Cancer Res 66:10815-10823.

Liu G, Yuan X, Zeng Z, Tunici P, Ng H, Abdulkadir IR, Lu L, Irvin D, Black KL, Yu JS. 2006b. Analysis of gene expression and chemoresistance of CD133+ cancer stem cells in glioblastoma. Mol Cancer 5:67.

Maher EA, Brennan C, Wen PY, Durso L, Ligon KL, Richardson A, Khatry D, Feng B, Sinha R, Louis DN and others. 2006. Marked genomic differences characterize primary and secondary glioblastoma subtypes and identify two distinct molecular and clinical secondary glioblastoma entities. Cancer Res 66:11502-11513.

Nagaraja GM, Kandpal RP. 2004. Chromosome 13q12 encoded Rho GTPase activating protein suppresses growth of breast carcinoma cells, and yeast two-hybrid screen shows its interaction with several proteins. Biochem Biophys Res Commun 313:654-665.

Ng DC-H, Chan S-F, Kok KH, Yam JWP, Ching Y-P, Ng IO-L, Jin D-Y. 2006. Mitochondrial targeting of growth suppressor protein DLC2 through the START domain. FEBS Lett 580:191-198.

Nigro JM, Misra A, Zhang L, Smirnov I, Colman H, Griffin C, Ozburn N, Chen M, Pan E, Koul D and others. 2005. Integrated array-comparative genomic hybridization and expression array profiles identify clinically relevant molecular subtypes of glioblastoma. Cancer Res 65:1678-1686.

Nutt CL, Mani DR, Betensky RA, Tamayo P, Cairncross JG, Ladd C, Pohl U, Hartmann C, McLaughlin ME, Batchelor TT and others. 2003. Gene expression-based classification of malignant gliomas correlates better with survival than histological classification. Cancer Res 63:1602-1607.

Okamoto OK, Oba-Shinjo SM, Lopes L, Marie SKN. 2007. Expression of HOXC9 and E2F2 are up-regulated in CD133(+) cells isolated from human astrocytomas and associate with transformation of human astrocytes. Biochim Biophys Acta 1769:437-442.

Phillips HS, Kharbanda S, Chen R, Forrest WF, Soriano RH, Wu TD, Misra A, Nigro JM, Colman H, Soroceanu L and others. 2006. Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. Cancer Cell 9:157-173.

Raghavan R, Steart PV, Weller RO. 1990. Cell proliferation patterns in the diagnosis of astrocytomas, anaplastic astrocytomas and glioblastoma multiforme: a Ki-67 study. Neuropathol Appl Neurobiol 16:123-133.

Sarkar S, Nuttall RK, Liu S, Edwards DR, Yong VW. 2006. Tenascin-C stimulates glioma cell invasion through matrix metalloproteinase-12. Cancer Res 66:11771-11780.

Schmid M, Sen M, Rosenbach MD, Carrera CJ, Friedman H, Carson DA. 2000. A methylthioadenosine phosphorylase (MTAP) fusion transcript identifies a new gene on chromosome 9p21 that is frequently deleted in cancer. Oncogene 19:5747-5754.

Shou J, Ali-Osman F, Multani AS, Pathak S, Fedi P, Srivenugopal KS. 2002. Human Dkk-1, a gene encoding a Wnt antagonist, responds to DNA damage and its overexpression sensitizes brain tumor cells to apoptosis following alkylation damage of DNA. Oncogene 21:878-889.

Smyth GK. 2005. Limma: linear models for microarray data. In Bioinformatics and Computational Biology Solutions using R and Bioconductor. Edited by Gentleman R, New York: Springer:397-420.

Soccio RE, Breslow JL. 2003. StAR-related lipid transfer (START) proteins: mediators of intracellular lipid metabolism. J Biol Chem 278:22183-22186.

Solomon DA, Kim JS, Jenkins S, Ressom H, Huang M, Coppa N, Mabanta L, Bigner D, Yan H, Jean W, Waldman T. 2008. Identification of p18INK4c as a Tumor Suppressor Gene in Glioblastoma Multiforme. Cancer Res 68:2564-2569

Stupp R, Pica A, Mirimanoff RO, Michielin O. 2007. A practical guide for the management of gliomas. Bul l Cancer 94:817–822.

Sun L, Hui A-M, Su Q, Vortmeyer A, Kotliarov Y, Pastorino S, Passaniti A, Menon J, Walling J, Bailey R and others. 2006. Neuronal and glioma-derived stem cell factor induces angiogenesis within the brain. Cancer Cell 9:287-300.

Tanwar MK, Gilbert MR, Holland EC. 2002. Gene expression microarray analysis reveals YKL-40 to be a potential serum marker for malignant character in human glioma. Cancer Res 62:4364-4368.

Tso C-L, Freije WA, Day A, Chen Z, Merriman B, Perlina A, Lee Y, Dia EQ, Yoshimoto K, Mischel PS and others. 2006. Distinct transcription profiles of primary and secondary glioblastoma subgroups. Cancer Res 66:159-167.

van den Boom J, Wolter M, Kuick R, Misek DE, Youkilis AS, Wechsler DS, Sommer C, Reifenberger G, Hanash SM. 2003. Characterization of gene expression profiles associated with glioma progression using oligonucleotide-based microarray analysis and real-time reverse transcription-polymerase chain reaction. Am J Pathol 163:1033-1043.

Wang C-C, Liao Y-P, Mischel PS, Iwamoto KS, Cacalano NA, McBride WH. 2006. HDJ-2 as a target for radiosensitization of glioblastoma multiforme cells by the farnesyltransferase inhibitor R115777 and the role of the p53/p21 pathway. Cancer Res 66:6756-6762.

Wonsey DR, Follettie MT. 2005. Loss of the forkhead transcription factor FoxM1 causes centrosome amplification and mitotic catastrophe. Cancer Res 65(12):5181-5189.

Ylagan LR, Quinn B. 1997. CD44 expression in astrocytic tumors. Mod Pathol 10:1239-1246.

Yuan BZ, Miller MJ, Keck CL, Zimonjic DB, Thorgeirsson SS, Popescu NC. 1998. Cloning, characterization, and chromosomal localization of a gene frequently deleted in human liver cancer (DLC-1) homologous to rat RhoGAP. Cancer Res 58:2196-2199.

FIGURES LEGENDS

Figure 1 — Outline of strategy to combine copy number alterations and gene expression data. The 'segmented' data set (a) and its paired gene expression data set (b), corresponding to the selected Minimal Common Regions (MCR), were used either for *Targeted study* (A) or *Correlation study* (B). For the *Targeted study*, statistical analysis (*t*-test) was performed and genes with BH *P*-values<0.05 and a concordant ratio of expression were selected. For the *Correlation study*, the Pearson correlation coefficient was calculated and the correlation matrix was generated. Each gene for which the diagonal value was up to 0.7 was selected. Graphical displays illustrate the possible case selections: (I) example of a gene under-expressed in a deleted region (selected in the *Targeted study*); (II) example of a gene gained in 3 glioblastomas with correlated over-expression in the corresponding tumors (selected in *the Correlation study*); (III) example of a gene significantly over-expressed with expressions levels correlated to the corresponding DNA dosage (selected in both studies).

Figure 2 — DNA Copy number alterations in Glioblastoma. A. Frequencies (%) of significant copy number alterations across the samples in segmented data (y axis) are plotted for each array CGH probe aligned along the x-axis in chromosomal order. Red=gains, green=losses, dark red and dark green indicate amplification and homozygous deletion, respectively. B. The result of the unsupervised hierarchical clustering of glioblastoma samples based on the altered minimal common regions (except chromosomes X and Y) is displayed. This clustering was done on 'segmented' data, using the Support Tree option in the TM4 microarray software suite (MeV), and

consensus clusters were built using bootstrapping with average linkage and Pearson correlation as distance metric. Probes are ordered according to their genomic position and chromosomes are noted below the heatmap. Colored tree branches denote the level of support resulting from the resampling (black: 100% support, yellow: 60-70% support). Transparent blue (subtype-1) and red (subtype-2) triangles were created to collect 2 principal groups of glioblastomas falling below terminal nodes in the tree using a fixed distance threshold.

Figure 3 — Targeted study: 261 genes with significant differential expression concordant to DNA copy number changes. The *Targeted study* identified 261 genes in minimal common regions with concordant and significant differential expression. A. CGH and expression profiles are plotted (right and left heatmap, respectively) on the y-axis by genomic map positions, samples are on the same order on the x-axis. Indications of minimal common regions are marked on the left. B. Results obtained by the combined functional annotation method developed by Aubry et al. (2006). The sources of annotation (goa: Gene Ontology Annotation database, pub: PubGene biomedical literature index and goapub: both) are provided and for each of them, the number of associated genes and *p*-values are reported. Groups of differentially expressed genes (up and down) are also indicated. C. Map of the top-scoring network from the functional network analysis is presented; nodes represent genes, with their shape representing the functional classes of the gene products, and edges indicate the biologic relationship between the nodes, which include physical and functional interactions. Associated functions are also mapped.

Figure 4 — Correlation study: 159 genes with expression highly correlated to DNA copy number changes. The Correlation study identified 159 genes in Minimal Common Regions with expression highly correlated with DNA copy number status (Pearson's correlation coefficient up to 0.7). A. CGH and expression profiles are plotted (right and left heatmap, respectively) on the y-axis by genomic map positions, samples are on the same order on the x-axis. Indications of minimal common regions are marked on the left. B. Results obtained by the combined functional annotation method developed by Aubry et al. (2006). The sources of annotation (goa: Gene Ontology Annotation database, pub: PubGene biomedical literature index and goapub: both) are provided and for each, the number of associated genes and p-value are reported. C. Map of the top-scoring network from the functional network analysis is presented; nodes represent genes, with their shape representing the functional classes of the gene products, and edges indicate the biologic relationship between the nodes, which include physical and functional interactions. Associated functions are also mapped.

Figure 5 — The 'cis-inducing DNA targeted genes' define a robust genomic signature grouping potential key initiating events in Glioblastoma. The '*cis*-inducing DNA targeted genes' signature, comprising the combined results from the *Targeted* and the *Correlated* studies and validated by data from the Sun et al. Study, is given. A. Expression profiles are plotted on the y-axis by genomic map positions and samples are ordered on the x-axis by their study of origin (blue box for GDS1962 data and pink box for our local data). Data values are expressed as log2 ratio between each glioblastoma and the mean of all the normal brain samples. B. and C. Maps of the first two top-scoring networks from the functional network analysis are presented; nodes represent genes,

with their shape representing the functional classes of the gene products, and edges indicate the biologic relationship between the nodes, which include physical and functional interactions. Associated functions are also mapped.

Figure 6 — Gene expression pattern subgrouping and evaluation of the homogeneous and heterogeneous part of the Glioblastoma 'cis-inducing DNA targeted genes' signature. Results of the unsupervised hierarchical clustering analyses of all glioblastomas (external data and Local data) based on the *Targeted Genes* (A) and on the *Correlated Genes* (B) are displayed. These analyses were performed using the Support Tree option in the TM4 microarray software suite (MeV), and consensus clusters were built using bootstrapping with complete linkage and Euclidian as distance metric. Probes are ordered according to their genomic position. Colored tree branches denote the level of support resulting from the resampling (black: 100% support and red-pink: no support). Transparent blue triangles were created to collect 2 principal groups of glioblastomas falling below terminal nodes in the tree using a fixed distance threshold.

Figure 7 — Validation of PCDH9 and STARD13 genes by RT Q- PCR.

PCDH9 and DLC2 were deemed particularly interesting because of their significant under-expression correlated to DNA alterations. Y-axis, mean relative level of mRNA (log2 based on the reference and commercial brain RNA as calibrator in glioblastoma (GBM) and non-neoplastic brain (N) samples); ** corresponds to $p < 0.01$.

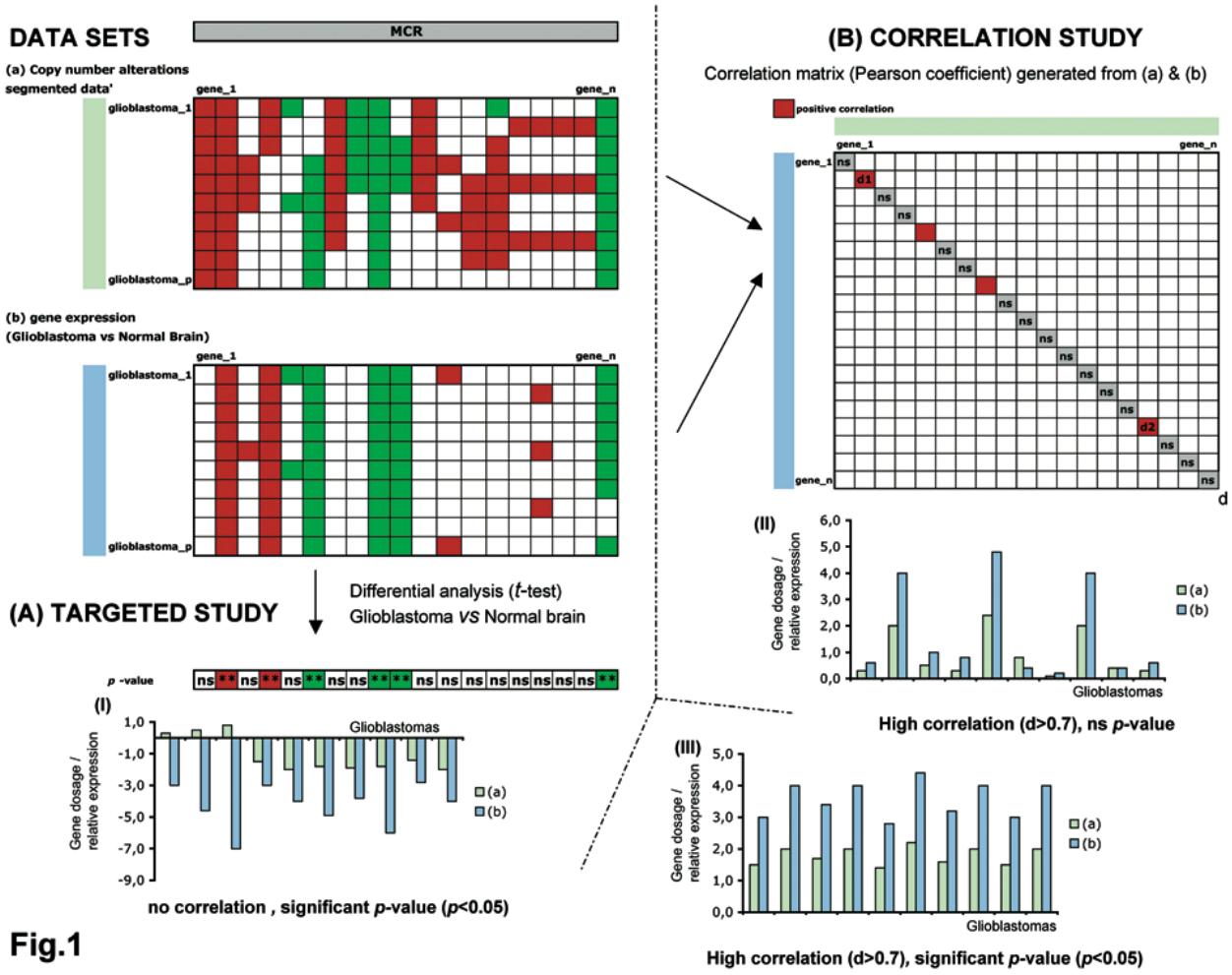

DATA SETS
MCR
(a) Copy number alterations segmented data'
gene_1
gene_n
glioblastoma_1
glioblastoma_p
(b) gene expression (Glioblastoma vs Normal Brain)
(A) TARGETED STUDY
Differential analysis (t-test)
Glioblastoma VS Normal brain
p -value
ns ** ns ** ns ** ns ns *** ns ns ns ns ns ns ns **
(I)
Gene dosage / relative expression
Glioblastomas
(a)
(b)
no correlation , significant p-value (p<0.05)
(B) CORRELATION STUDY
Correlation matrix (Pearson coefficient) generated from (a) & (b)
positive correlation
d1
d2
d
(II)
High correlation (d>0.7), ns p-value
(III)
High correlation (d>0.7), significant p-value (p<0.05)

Fig.1

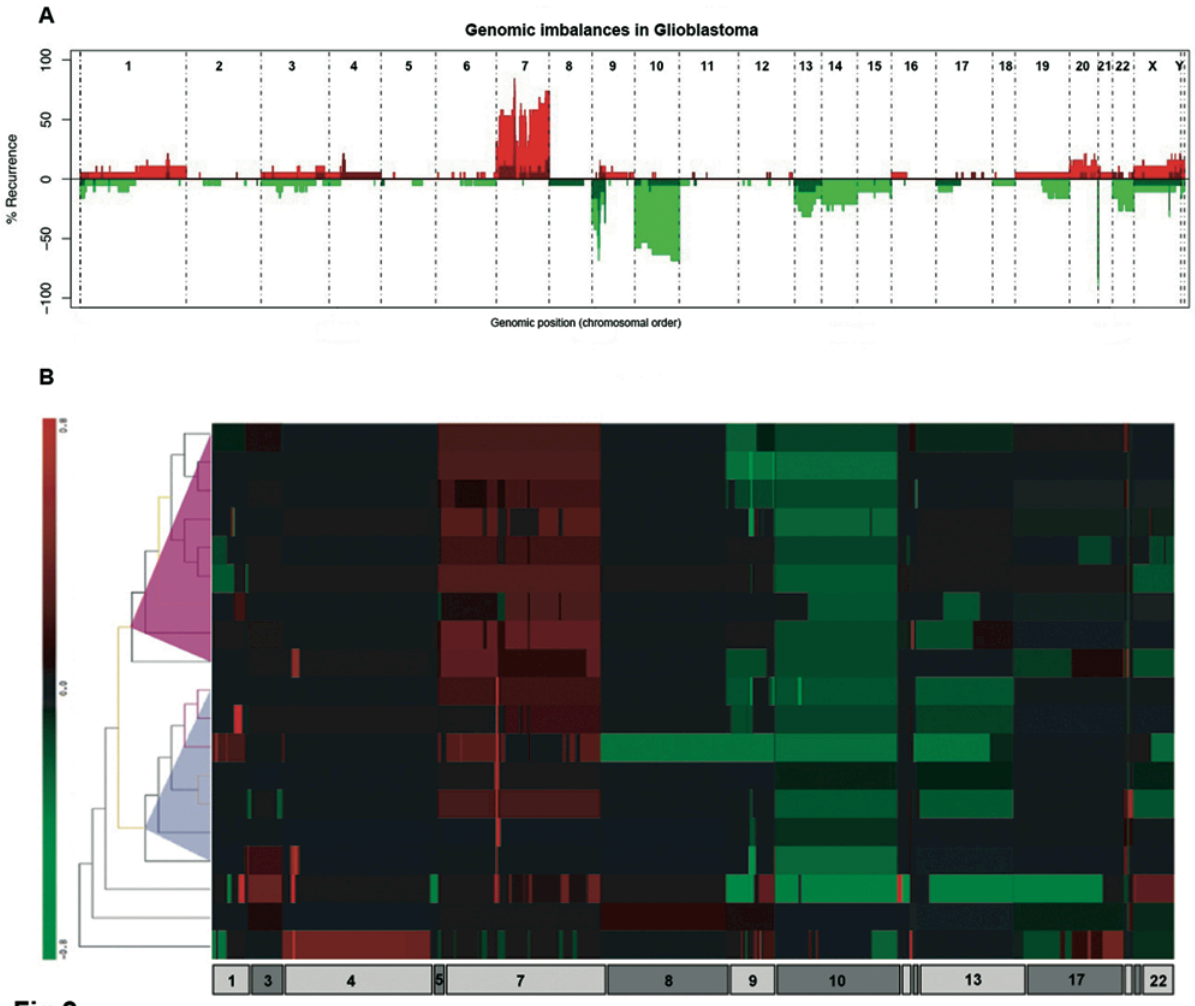
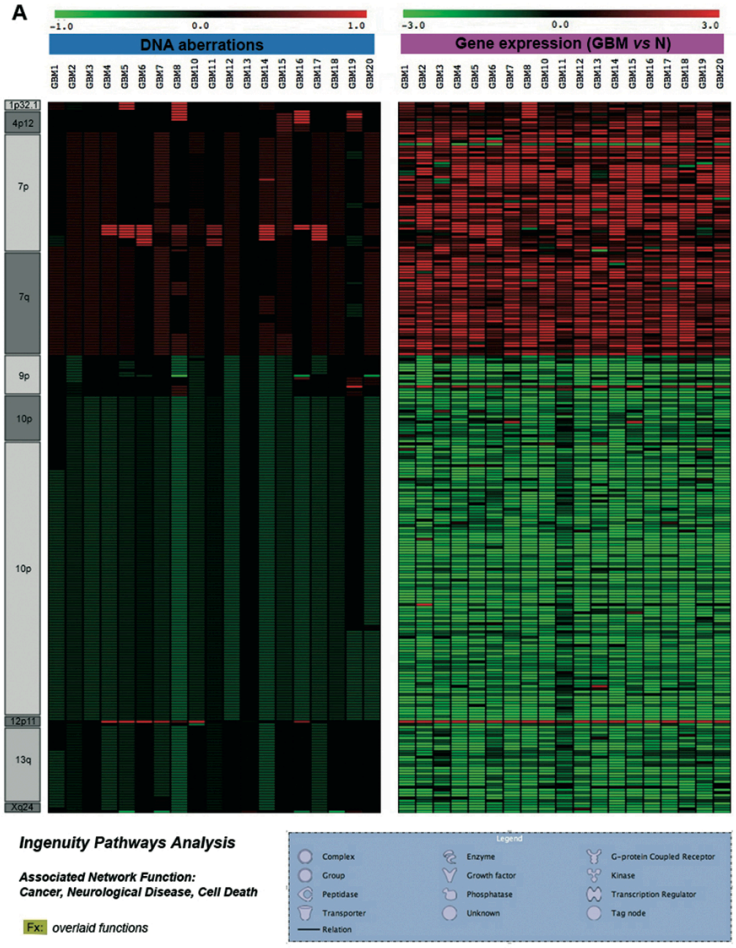

A
Genomic imbalances in Glioblastoma
% Recurrence
100
50
0
−50
−100
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Y
Genomic position (chromosomal order)
B
1
3
4
5
7
8
9
10
13
17
22

Fig.2

A



Ingenuity Pathways Analysis

Associated Network Function:
Cancer, Neurological Disease, Cell Death

Fx: *overlaid functions*

Fig.3

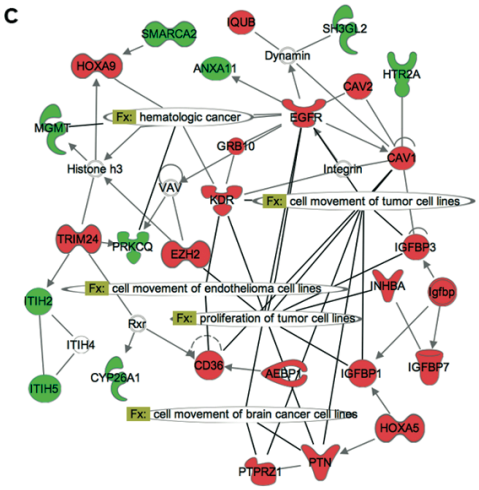
B

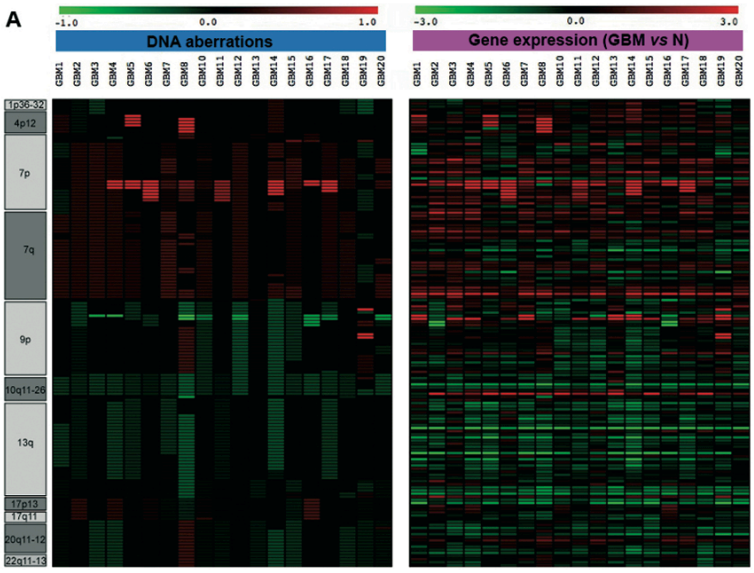
GO ID	GO Term	Number of gene (goa)(pub)	Annotation	*p* -value
GO:0007049	cell cycle	(3),(10)	goapub	4.75e-12
GO:0000074	regulation of progression through cell cycle	(1),(22)	goapub	2.12e-05
GO:0007050	cell cycle arrest	7	pub	2.92e-05
GO:0007049	cell cycle	(1),(8)	goapub	4.75e-12
GO:0007052	mitotic spindle organization and biogenesis	3	pub	1.24e-01
GO:0006915	apoptosis	6	pub	8.69e-27
GO:0006917	induction of apoptosis	(1),(6)	goapub	3.28e-06
GO:0006919	caspase activation	2	pub	5.34e-05
GO:0006915	apoptosis	(1),(4)	goapub	1.01e-18
GO:0042981	regulation of apoptosis	(2),(2)	goapub	2.10e-04
GO:0006309	DNA fragmentation during apoptosis	4	pub	2.44e-03
GO:0007155	cell adhesion	(3),(9)	goapub	9.97e-11
GO:0016337	cell-cell adhesion	7	pub	2.13e-02
GO:0016339	calcium-dependent cell-cell adhesion	3	pub	1.79e-02
GO:0007157	heterophilic cell adhesion	5	pub	5.34e-06
GO:0007155	cell adhesion	(2),(6)	goapub	8.12e-08
GO:0007157	heterophilic cell adhesion	9	pub	3.31e-02
GO:0006397	mRNA processing	(1),(4)	goapub	5.07e-03
GO:0000245	spliceosome assembly	2	pub	9.93e-02
GO:0000398	nuclear mRNA splicing, via spliceosome	12	pub	1.11e-02
GO:0006281	DNA repair	(2),(6)	goapub	1.61e-04
GO:0006302	double-strand break repair	3	pub	2.26e-01
GO:0007165	signal transduction	(11),(17)	goapub	1.72e-12
GO:0007265	Ras protein signal transduction	2	goa	2.88e-02
GO:0007243	protein kinase cascade	6	pub	2.35e-02
GO:0007259	JAK-STAT cascade	2	pub	3.36e-02
GO:0001525	angiogenesis	(1),(5)	goapub	2.00e-04
GO:0045765	regulation of angiogenesis	5	pub	1.48e-01
GO:0045766	positive regulation of angiogenesis	2	pub	1.47e-03

under-expressed genes

over-expressed genes

C





B

GO ID	GO Term	Number of gene (goa)(pub)	Annotation	*p* -value
GO:0007049	cell cycle	(9),(23)	goapub	3.44e-09
GO:0000074	regulation of progression through cell cycle	(3),(31)	goapub	8.50e-03
GO:0000075	cell cycle checkpoint	(1)(7)	goapub	6.02e-04
GO:0007050	cell cycle arrest	(2)(6)	goapub	7.01e-03
GO:0007067	mitosis	(3),(9)	goapub	2.28e-05
GO:0045786	negative regulation of progression through cell cycle	(1)(3)	goapub	1.42e-03
GO:0012501	programmed cell death	4	pub	5.78e-05
GO:0007155	**cell adhesion**	**(2),(9)**	**goapub**	**1.37e-10**
GO:0008283	**cell proliferation**	**2**	**goa**	**6.56e-02**
GO:0008380	RNA splicing	(2),(6)	goapub	1.65e-03
GO:0000245	spliceosome assembly	(2),(1)	goapub	4.20e-03
GO:0000398	nuclear mRNA splicing, via spliceosome	(2),(11)	goapub	6.52e-03
GO:0006281	**DNA repair**	**(2),(9)**	**goapub**	**7.80e-06**
GO:0006310	**DNA recombination**	**(2),(4)**	**goapub**	**1.08e-03**
GO:0000731	**DNA synthesis during DNA repair**	**3**	**pub**	**8.03e-02**
GO:0006302	**double-strand break repair**	**3**	**pub**	**3.94e-02**
GO:0000012	**single strand break repair**	**(2)(2)**	**goapub**	**1.13e-03**
GO:0007399	nervous system development	(2),(3)	goapub	3.21e-04
GO:0007420	brain development	(1),(7)	goapub	2.65e-06
GO:0000079	**regulation of cyclin-dependent protein kinase activity**	**(2)(2)**	**goapub**	**1.30e-03**

C

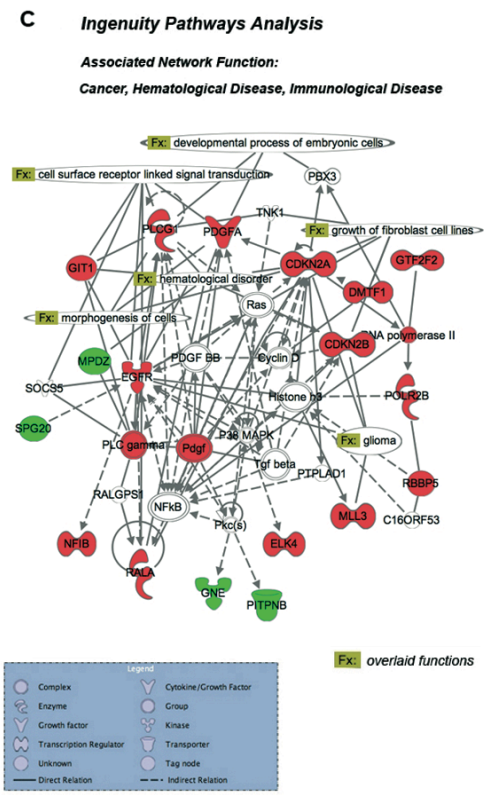


Fig.4

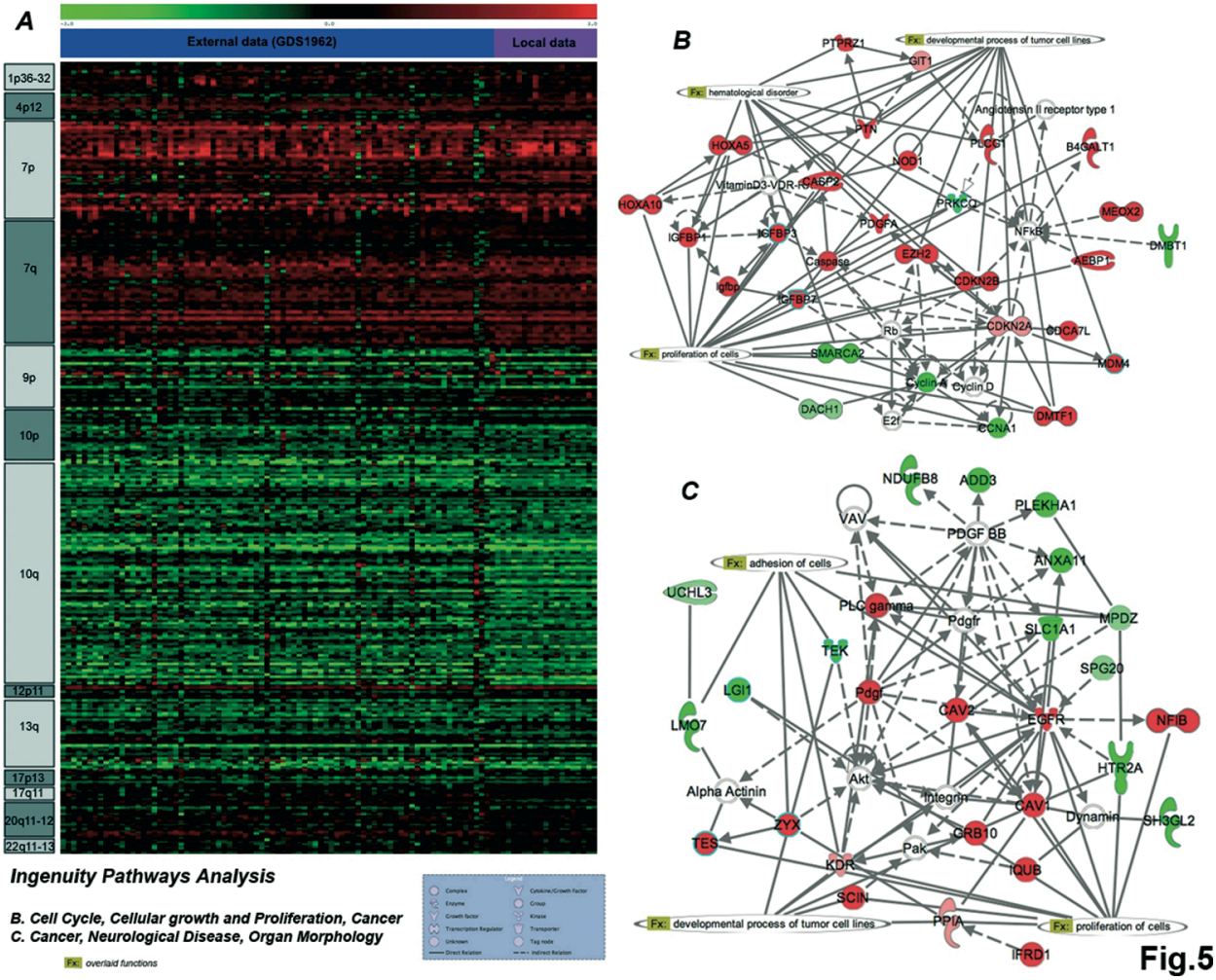
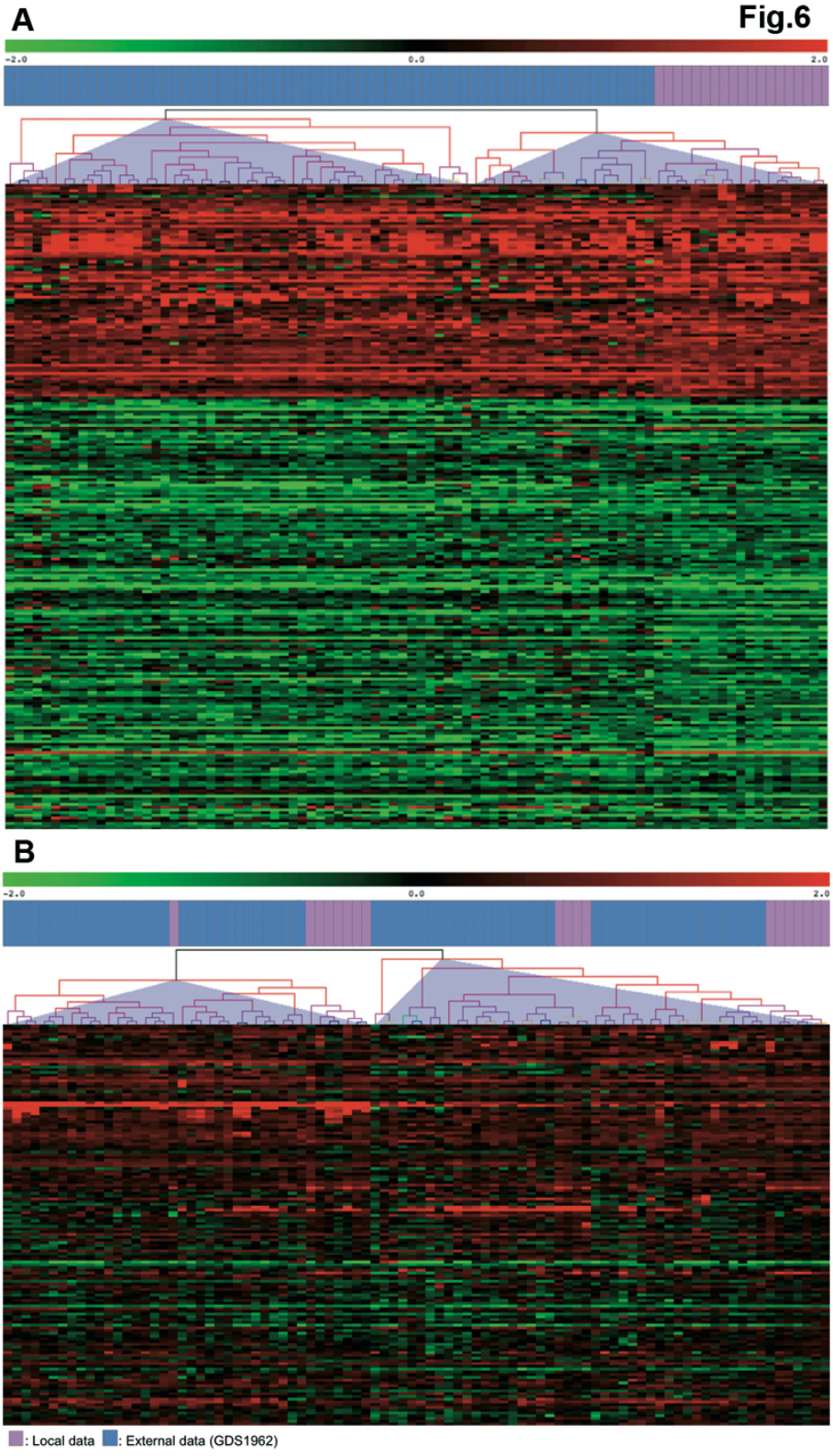

A
External data (GDS1962)
Local data
1p36-32
4p12
7p
7q
9p
10p
10q
12p11
13q
17p13
17q11
20q11-12
22q11-13
B
PTPRZ1
GIT1
Fx: developmental process of tumor cell lines
Fx: hematological disorder
Angiotensin II receptor type 1
PTN
HOXA5
PLCG
B4GALT1
NOD1
VitaminD3-VDR-RXR
CA6P2
HOXA10
PRKCQ
MEOX2
IGFBP4
IGFBP3
PDGFA
NFkB
DMBT1
Caspase
EZH2
4EBP1
Igfbp
CDKN2B
IGFBP2
Rb
CDKN2A
CDCA7L
Fx: proliferation of cells
SMARCA2
MDM4
Cyclin A
Cyclin D
DACH1
E2f
DMTF1
CCNA1
C
NDUFB8
ADD3
PLEKHA1
VAV
PDGF BB
ANXA11
Fx: adhesion of cells
UCHL3
PLC gamma
Pdgfr
SLC1A1
MPDZ
TEK
SPG20
LGI1
Pdgf
CAV2
EGFR
NFIB
LMO7
HTR2A
Akt
Alpha Actinin
Integrin
CAV1
Dynamin
SH3GL2
ZYX
CRB10
TES
Pak
KDR
IQUB
SCIN
Fx: developmental process of tumor cell lines
PPIA
Fx: proliferation of cells
IFRD1
Ingenuity Pathways Analysis
B. Cell Cycle, Cellular growth and Proliferation, Cancer
C. Cancer, Neurological Disease, Organ Morphology
Fx: overlaid functions

Fig.5

Fig.6



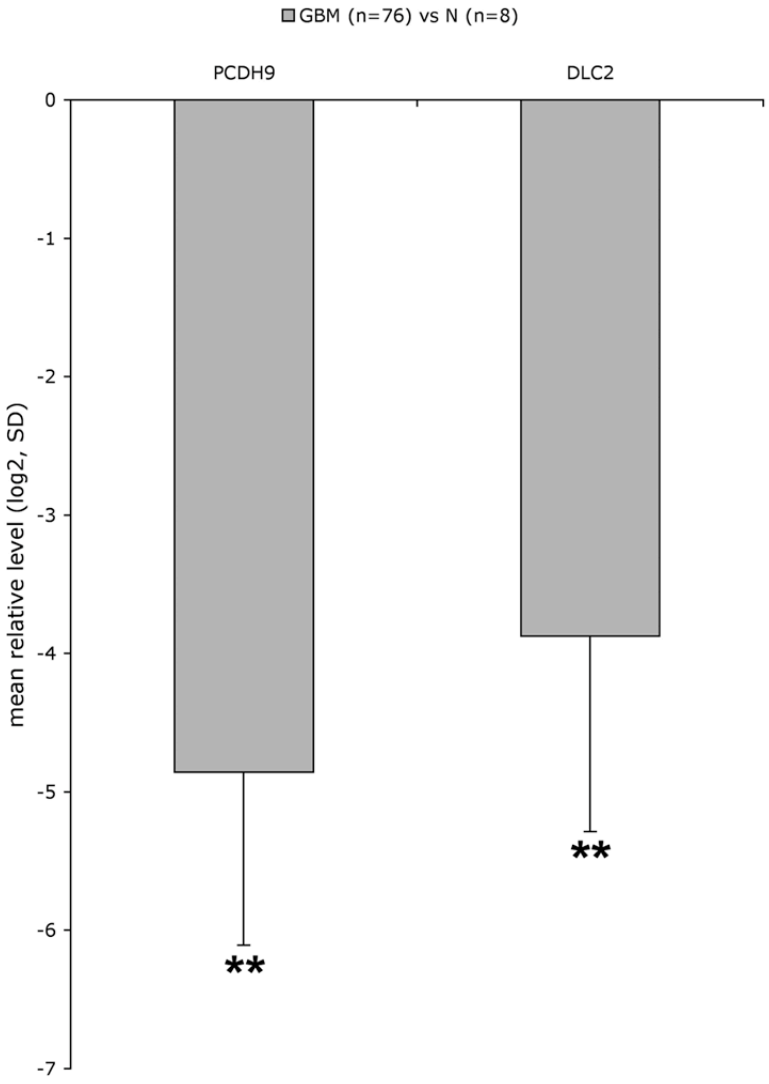

GBM (n=76) vs N (n=8)
PCDH9
DLC2
mean relative level (log2, SD)
0
-1
-2
-3
-4
-5
-6
-7
**
**

Fig 7.